Energy-driven Drag at Charge Neutrality in Graphene

Justin C. W. Song[1,2] and Leonid S. Levitov[1]

[1] *Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA and*
[2] *School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*

Coulomb coupling between proximal layers in graphene heterostructures results in efficient energy transfer between the layers. We predict that, in the presence of correlated density inhomogeneities in the layers, vertical energy transfer has a strong impact on lateral charge transport. In particular, for Coulomb drag it dominates over the conventional momentum drag near zero doping. The dependence on doping and temperature, which is different for the two drag mechanisms, can be used to separate these mechanisms in experiment. We predict distinct features such as a peak at zero doping and a multiple sign reversal, which provide diagnostics for this new drag mechanism.

Recently developed vertical heterostructures[1] comprised of a few graphene layers separated by an atomically thin insulating layer afford new ways to probe the effects of electron interactions at the nanoscale. Typical layer separation in these structures (1-2 nm) can be very small compared to the characteristic electron lengthscales such as the de Broglie wavelength and the screening length. This defines a new strong-coupling regime wherein the interlayer and intralayer interactions are almost equally strong. Fast momentum transfer between electron subsystems in the two layers and strong Coulomb drag have been predicted in this regime[2–7] with characteristic dependence on doping, temperature and layer separation distinct from that in previously studied systems[8]. Recently, measurements of strong drag in graphene were reported[9, 10].

In this article we focus on another effect that becomes prominent in the strong coupling regime: vertical energy transfer mediated by interlayer electron-electron scattering. We predict that this process can give rise to lateral energy flow in the electron system, which under the conditions discussed below can directly impact electric transport. In particular, it leads to a characteristic contribution to the in-plane resistivity and dominates drag near charge neutrality (CN), see Fig.1 and Fig.3.

Interlayer electron-electron scattering, which governs drag, is strongest at CN since the long-range Coulomb interactions become unscreened near CN. The scattering rate decreases with doping away from CN as $\gamma \propto \nu^{-1}(\mu)$, where $\nu(\mu)$ is the massless Dirac density of states, $\mu \gg k_B T$ is the chemical potential, see Eq.(8). The $1/\nu$ scaling is completely analogous to that found for intralayer scattering [11, 12]. Crucially, while both the energy and momentum transfer rates peak at CN, their impact on drag is markedly different. Since the sign of momentum drag depends on the polarity of charge carriers[13], momentum drag vanishes at CN [2–7]. In contrast, energy-driven drag features a peak at CN, see Fig.3.

Our drag mechanism arises due to coupling between vertical energy transfer and lateral charge and energy transport via spatial density inhomogeneity which is intrinsic to graphene. Density inhomogeneity is known to
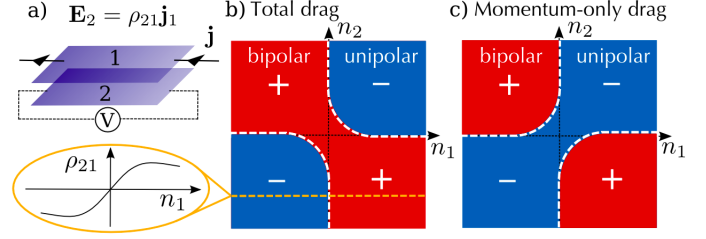

FIG. 1: Different mechanisms for Coulomb drag in graphene heterostructures. Energy-driven drag dominates over momentum drag near zero doping, whereas momentum drag dominates at higher doping. The sign of the drag response depends on carrier polarity (a). For potential fluctuations of equal sign in the two layers, Eq.(1), the net drag (b) features a pair of nodal lines with an avoided crossing. The drag sign at zero doping is positive, as in the bipolar regime. This is distinct from momentum-only contribution to drag (c) smeared by correlated density fuctuations, $\delta\mu_1 \approx \delta\mu_2$.

be particularly strong at CN in the electron-hole puddle regime[14], providing the dominant disorder potential in clean samples. When a charge current is applied in layer 1, density inhomogeneity produces spatially varying heating/cooling [see Eq.(2)]. Strong thermal coupling between the electron systems in the two layers, mediated by the interlayer energy transfer, leads to a temperature pattern in layer 2 that tracks that in layer 1, $\delta T_2(r) \approx \delta T_1(r)$. Further, since the disorder correlation length ξ_{dis} can reach 100 nm in G/BN heterostructures[15, 16], exceeding the layer separation by orders of magnitude, the potential fluctuations are nearly identical in the two layers,

$$\langle \delta\mu_1(r)\delta\mu_2(r') \rangle > 0 \tag{1}$$

for $r \approx r'$. As a result, the position-dependent thermopower induced by the gradient $\nabla\delta T_2(r)$ is correlated with the heating/cooling pattern in layer 1, giving rise to a nonzero ensemble-averaged drag voltage in layer 2.

Our mechanism predicts a particular sign of the energy contribution to drag. As a result, the density dependence for the net drag (energy and momentum combined) features a split-up pattern of nodal lines with an "avoided crossing" at zero doping, as illustrated in Fig. 1 (b). The double sign change along the main diagonal $n_1 = n_2$ and
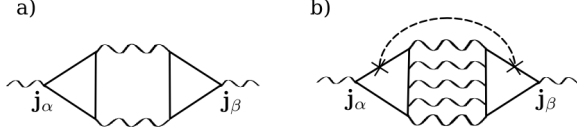
FIG. 2: Feynman diagrams for momentum (a) and energy (b) contributions to drag. Wavy lines represent interactions, dashed line represents disorder averaging. The ladder in (b) represents a long-wavelength charge-neutral mode.

the peak at $n_{1,2} = 0$ make the energy-driven drag easy to distinguish experimentally.

As a parenthetical remark, the correlated density inhomogeneity, Eq.(1), also affects the momentum drag, however its effect is opposite to that of the energy contribution (see Fig.1 b and c). If momentum drag were the dominant contribution near zero doping, the pattern of nodal lines would be such that the drag sign was constant along the main diagonal. This qualitative difference makes it easy to differentiate between the two cases.

Below we develop a hydrodynamic framework to describe energy-driven drag. The neutral modes (particle-hole excitations, or temperature imbalance) which mediate drag in our mechanism are of a long-range character, propagating over distances much larger than the inelastic mean free path $\ell = v/\gamma$. The length scales relevant for our drag mechanism are the electron-lattice cooling length and the inhomogeneity correlation length $\xi_T, \xi_{\text{dis}} \gg \ell$ whichever is the smallest (ξ_T can be as large as several microns, even at room temperature [17–19]).

It is instructive to compare the Feynman diagrams describing the conventional momentum drag and our mechanism (see Fig.2). While the characteristic momenta are large ($\sim k_F$) for the former contribution (Fig.2a), the latter contribution (Fig.2b) includes ladder diagrams representing neutral modes propagating over distances of order ξ_T. As a result, energy-driven drag is captured by a hydrodynamic framework which involves charge current $\mathbf{j}$ and heat current $\mathbf{j}_q$, which in the ballistic transport regime are related by

$$\mathbf{j}_q(\mathbf{r}) = Q(n)\mathbf{j}, \quad Q[n(\mathbf{r})] = \mathcal{S}[n(\mathbf{r})]T/e, \quad (2)$$

where $\mathcal{S}(n)$ is the entropy per carrier, $n(\mathbf{r})$ is the density profile, and $e < 0$ is carrier charge. In the ballistic regime, using electron temperature approximation, we find[20]

$$Q = \frac{2\pi^2 k_B^2 T^2 \mu}{3e(\mu^2 + \Delta^2(T))}. \quad (3)$$

where $\Delta(T)$ accounts for the Dirac point broadening due to disorder and thermal fluctuations.

To illustrate the relation between energy and charge transport, we first analyze in-plane resistivity in a single layer. According to Eq.(2), spatial inhomogeneity leads to heating/cooling in the presence of uniform charge current (a la the Joule-Thomson process). The spatial temperature profile can be found from $-\nabla \kappa \nabla \delta T +$ $\lambda \delta T = -\nabla \cdot \mathbf{j}_q$, where κ is the thermal conductivity and $\lambda \delta T$ is the electron-lattice cooling power. A temperature gradient $\nabla \delta T$ drives thermopower, providing additional dissipation and thereby increasing resistivity. Onsager reciprocity combined with Eq.(2) gives $\mathbf{E}(r) = -(Q[n(\mathbf{r})]/T)\nabla \delta T$[20]. Taking an ensemble average over small density fluctuations, $\delta \mu \ll k_B T, \mu$, we find an increase in the in-plane resistivity, $\langle \rho_{\alpha\beta} \rangle = \rho_{\alpha\beta}^0 + \Delta\rho_{\alpha\beta}$, $[\alpha(\beta) = x, y]$, where

$$\Delta\rho_{\alpha\beta} = \frac{1}{T} \sum_{|\mathbf{q}| \lesssim 1/\ell} \frac{\langle \delta Q(-\mathbf{q})\delta Q(\mathbf{q}) \rangle}{\kappa q^2 + \lambda} q_\alpha q_\beta. \quad (4)$$

Since the derivative $\partial Q/\partial \mu$ peaks at $\mu = 0$, this results in $\Delta\rho_{\alpha\beta}$ that peaks at CN. The temperature dependence estimated below is $\Delta\rho \propto T^2$, reminiscent of super-linear power laws for resistivity frequently observed at small doping[21]. A contribution of nonthermal modes to $\Delta\rho$ was analyzed in Ref.[22].

Generalizing this analysis to two layers coupled by vertical energy transfer and accounting for correlated density fluctuations, Eq.(1), we find an ensemble-averaged drag response $\mathbf{E}_2 = \rho_{21}\mathbf{j}_1$,

$$\rho_{21} = \frac{1}{2T\tilde{\kappa}} \frac{\partial Q}{\partial \mu_1} \frac{\partial Q}{\partial \mu_2} \sum_{\mathbf{q}} \frac{\langle \delta\mu_2(-\mathbf{q})\delta\mu_1(\mathbf{q}) \rangle}{1 + \xi_c^2 \mathbf{q}^2}, \quad (5)$$

where $\tilde{\kappa} = \kappa_1 + \kappa_2$ is the net thermal conductivity of the two layers, μ is the chemical potential, $\xi_c \sim \ell$ is the interlayer cooling calculated below.

Because the sign of the correlator in Eq.(1) is positive, energy-driven drag has the same sign as $\Delta\rho_{\alpha\beta}$ in Eq.(4), i.e. is *positive at zero doping*. This results in a double sign change along the main diagonal $n_1 = n_2$, as pictured in Fig. 1 (b). The density dependence for ρ_{21} features a peak at zero doping (see Fig.3) which is a hallmark of the energy-driven drag regime.

We note that, if the sign of the correlator $\langle \delta\mu_1 \delta\mu_2 \rangle$ were negative, as expected for strain-induced charge puddles[23], our analysis carries through but predicts a negative drag at zero doping. Hence drag is a useful tool for probing the origin of inhomogeneity in graphene.

We begin by studying the energy transfer between the electronic systems in the two layers (Fig.1(a)). This is described by the Hamiltonian

$$\mathcal{H} = \sum_i \int d^2\mathbf{r} \psi_i^\dagger(\mathbf{r}) \Big[-i\hbar v \sigma \cdot \nabla + \delta\mu_i(\mathbf{r}) \Big] \psi_i(\mathbf{r}) + \mathcal{H}_{\text{el-el}} \quad (6)$$

where $\{i, j\} = 1...2N$ index layer, and spin/valley degrees of freedom, $\delta\mu(\mathbf{r})$ describes the slowly varying disorder potential, v is the Fermi velocity, and the electron-electron interactions are $\mathcal{H}_{\text{el-el}} = \sum_{\mathbf{q},\mathbf{k},\mathbf{k}',i,j} V_{ij}(\mathbf{q})\psi_{\mathbf{k}+\mathbf{q},i}^\dagger \psi_{\mathbf{k}'-\mathbf{q},j}^\dagger \psi_{\mathbf{k}',j} \psi_{\mathbf{k},i}$ with $V_{ij}(\mathbf{q})$ the interaction.

In our analysis, we ignore the correction due to finite layer separation d, approximating the interlayer interaction by the bare Coulomb interaction, $V_{ij}(\mathbf{q}) =$

$V^0_{\mathbf{q}}e^{-d|\mathbf{q}|} \approx V^0_{\mathbf{q}} = 2\pi e^2/\varepsilon|\mathbf{q}|$ with ε the background dielectric constant. This approximation is valid when the lengthscale d is small compared to the screening length and Fermi wavelength in the layers, which is the case for systems of current interest[1]. The random-phase approximation then yields a screened interaction $V_{ij}(\mathbf{q}) = V^0_{\mathbf{q}}/[1 - V^0_{\mathbf{q}}(\Pi_1(\mathbf{q},\omega) + \Pi_2(\mathbf{q},\omega))]$ for i,j in different layers.

We describe the energy distribution of carriers in each layer by a Fermi distribution at temperatures $T_{1,2}$. Using Fermi's golden rule we can calculate the rate of energy exchange between the two layers (see Appendix). In the degenerate limit $\mu_1, \mu_2 \gg k_B T$, we obtain the energy loss power for layer 1 as

$$\mathcal{J}_{12} = \frac{6\zeta(4)}{\hbar^3 v^2} \frac{\nu_1 \nu_2 k_B^4}{(\nu_1 + \nu_2)^2}\left(T_1^4 \ln\frac{T_0}{T_1} - T_2^4 \ln\frac{T_0}{T_2}\right) \quad (7)$$

where $\nu(\mu)$ is the total density of states in each layer, and $k_B T_0 = v(2\pi e^2/\varepsilon)(\nu_1 + \nu_2)$. Notably, for equal densities $\mathcal{J}_{12}$ does not depend on the Fermi surface size. For equal densities and small temperature differences between the layers $T_1 \approx T_2$, we obtain the cooling rate

$$\gamma = \frac{1}{C_{\rm el}}\frac{d\mathcal{J}_{12}}{dT} = \frac{9\zeta(4)k_B^2 T^2}{\pi\mu\hbar}\ln\frac{T_0}{T} \quad (8)$$

where the heat capacity $C_{\rm el} = \pi^2/3 k_B^2 T \nu(\mu)$ and the density of states $\nu(\mu) = 2\mu/(\pi\hbar^2 v^2)$ for the degenerate limit have been used. The rate γ increases as μ goes towards neutrality, but is already quite large for μ away from neutrality. For typical values $\mu = 100\,\mathrm{meV}$, $T = 300\,\mathrm{K}$, the rate γ is about $29\,\mathrm{ps}^{-1}$, orders of magnitude faster than the electron-lattice cooling rates[17–19].

Vertical energy transfer couples heat transport in the two layers, so that the layer temperatures T_1, T_2 obey

$$-\nabla\kappa_1\nabla\delta T_1 + a(\delta T_1 - \delta T_2) + \lambda\delta T_1 = -\nabla\cdot\mathbf{j}_{q,1}$$
$$-\nabla\kappa_2\nabla\delta T_2 + a(\delta T_2 - \delta T_1) + \lambda\delta T_2 = 0 \quad (9)$$

where $a = d\mathcal{J}_{12}/dT$ [see Eq.(7)] and λ describes electron-lattice cooling. We consider only a response linear in the applied current, $\mathbf{j}$, neglecting the quadratic joule heating term. Inverting the coupled linear equations, we find an increase in temperature in layer 2, $\delta T_2(\mathbf{r})$, that is driven by current in layer 1 as

$$\delta T_2(\mathbf{r}) = -\frac{a}{\widehat{L}_1\widehat{L}_2 - a^2}(\mathbf{j}_1 \cdot \nabla)Q[n_1(\mathbf{r}), T], \quad (10)$$

where $\mathbf{j}_{q,1}$ is the heat current, Eq.(2), where $\widehat{L}_i = -\nabla\kappa_i\nabla + a + \lambda$. In what follows we suppress the λ term since electron-lattice cooling is slow. Eq.(10) then predicts a value for the interlayer cooling length $\xi_c = \sqrt{\kappa_1\kappa_2/[(\kappa_1 + \kappa_2)a]}$. The induced temperature profile, $\delta T_2(\mathbf{r})$, creates thermal gradients that can drive a local thermopower via $\mathbf{E}_2(r) = -(Q[n_2(\mathbf{r})]/T)\nabla\delta T_2$.

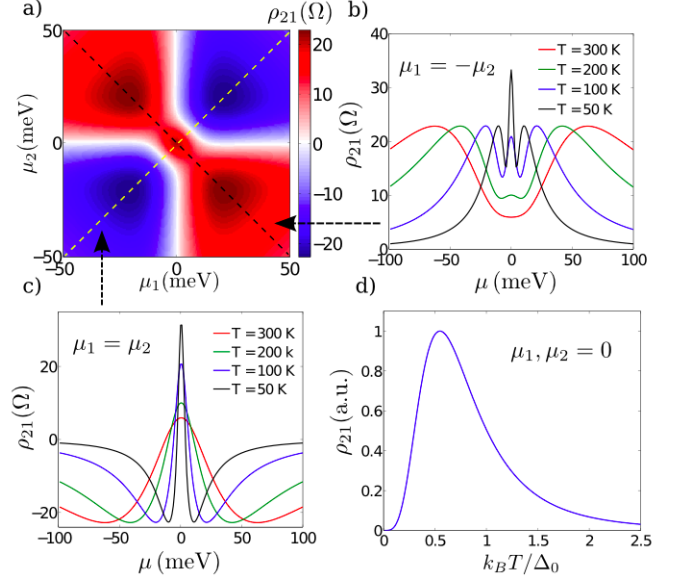FIG. 3: (a) Total drag resistivity $\rho_{21}^{\rm (tot)} = \rho_{21}^{\rm (m)} + \rho_{21}^{\rm (e)}$ vs. chemical potentials in the two layers, evaluated from Eq.(11) and Eq.(5) at $T = 100\,\mathrm{K}$, producing a peak at $\mu_{1,2} = 0$ (see text for parameter values used). (b,c) Slices $\mu_1 = \mu_2$ and $\mu_1 = -\mu_2$ at different temperatures. Note a three-peak structure in slice (b) and two sign changes close to CN in (c). (d) Temperature dependence of the peak at $\mu_{1,2} = 0$.

Spatial fluctuations in thermopower are governed by density fluctuations via Eq.(10). In particular, close to neutrality the local thermopower will exhibit regions of both positive and negative sign, leading to a spatial pattern of the drag resistivity. As discussed above, the correlations between $\delta\mu_1$ and $\delta\mu_2$, Eq.(1), lead to a nonzero ensemble-averaged drag resistivity. In the limit $\delta\mu_{1,2} \ll k_B T, \mu_{1,2}$ we write $Q_i(\mathbf{r}) = \langle Q_i(\mathbf{r})\rangle + \frac{\partial Q}{\partial\mu_i}\delta\mu_i(\mathbf{r})$. Passing to Fourier harmonics via $\langle\delta\mu_1(\mathbf{r})\delta\mu_2(\mathbf{r}')\rangle = \sum_{\mathbf{q}} e^{i\mathbf{q}(\mathbf{r}-\mathbf{r}')}\langle\delta\mu_1(-\mathbf{q})\delta\mu_2(\mathbf{q})\rangle$, we obtain Eq.(5).

The fact that fluctuating local thermopower, exhibiting both positive and negative signs, does not average to zero is surprising. This happens because the inhomogeneity in heat current and thermopower arise from the same source: electron-hole puddles. Energy-driven drag resembles mutual drag described by Laikhtman and Solomon [24] in semiconducting heterostructures where doping at contacts produced a similar correlation between Peltier heating/cooling and thermopower. Energy-driven drag in Graphene differs from Ref.[24] in that density inhomogeneity is intrinsic, occurs throughout the sample (not just at the contacts), and on a far smaller scale.

To see how energy-driven drag, Eq.(5), affects the total experimentally measured drag we need to account for momentum drag, $\rho_{21}^{\rm (m)}$. We use a model that captures the main qualitative features of momentum drag:

$$\rho_{21}^{\rm (m)} = \tilde{\rho}_{21}^{\rm (m)}\frac{h}{e^2}(k_B T)^2\frac{\mu_1}{(\mu_1^2 + \eta k_B^2 T^2)}\frac{\mu_2}{(\mu_2^2 + \eta k_B^2 T^2)}. \quad (11)$$

This expression, with the prefactor $\tilde{\rho}_{21}^{(m)} = -1.4\alpha^2/(2\pi\eta^2)$, the effective interaction strength $\alpha = 0.05$, and $\eta = 6.25$, was obtained by fitting the functional dependence derived in Ref. [7] in the doping region $-10 < \mu/k_BT < 10$.

Combining this with $\rho_{21}^{(e)}$ in Eq.(5), we obtain the total drag $\rho_{21}^{(tot)} = \rho_{21}^{(m)} + \rho_{21}^{(e)}$ plotted in Fig.3. Here we have used $\kappa = (\mu^2 + \Delta^2(T))/\hbar T$ [25], Eq.(2) and assumed Gaussian correlations with average square density fluctuations $\langle \delta\mu^2 \rangle \approx 25\,\text{meV}^2$ and $\xi_{\text{dis}} = 100\,\text{nm}$ [15, 16]. We note that the functional form of the correlator in Eq.(1) does not impact the qualitative behavior. The obtained values of total drag are compatible with measured drag resistivities reported in Refs.[9, 10].

The density dependence of total drag plotted in Fig.3 (a) can be used to distinguish the two drag mechanisms in experiments. Namely, the peak at zero doping is due to energy-driven drag. On the slice $\mu_1 = -\mu_2$ (black dashed line) this peak is surrounded by two peaks dominated by the momentum contribution [Fig.3(b)]. On the slice $\mu_1 = \mu_2$ (yellow dashed line) the two mechanisms produce contributions of opposite sign, resulting in a double sign change [Fig.3(c)]. This provides a clear means of discerning the energy-driven regime.

The temperature dependence can be estimated as follows. At very low T such that $\ell, \xi_c \gtrsim \xi_{\text{dis}}$, the sum in Eq.(5) is cut at $1/\ell$, giving $\rho_{21} \propto T^8$. At not too low T such that $\xi_c, \ell \lesssim \xi_{\text{dis}}$, the sum in Eq.(5) yields $\sum_{\mathbf{q}} \langle \delta\mu_1(-\mathbf{q})\delta\mu_2(\mathbf{q}) \rangle = \langle \delta\mu_1(\mathbf{r})\delta\mu_2(\mathbf{r}') \rangle_{\mathbf{r}=\mathbf{r}'}$. This gives a non-monotonic T dependence [see Fig. 3 (d)]

$$\rho_{21} \propto \frac{T^4}{\left(\Delta_0^2 + \eta(k_BT)^2\right)^3} \langle \delta\mu_1(\mathbf{r})\delta\mu_2(\mathbf{r}') \rangle_{\mathbf{r}=\mathbf{r}'}. \qquad (12)$$

A similar non-monotonic T dependence arises for in-plane resistivity $\Delta\rho_{\alpha\beta}$. Interestingly, a peak in drag resistance at CN with non-monotonic temperature dependence was recently reported in Ref.[10].

The above analysis can be easily extended to describe the diffusive limit where the elastic mean free path is shorter than the inelastic mean free path, $\ell' < \ell$. Our hydrodynamic approach remains valid in this regime, with the quantity $Q = sT$ where s is the Seebeck coefficient. The energy-driven drag is still given by Eq.(5), with s and κ described by the Mott and Wiedemann-Franz relations:

$$s = \frac{\pi^2}{3e} k_B^2 T \frac{\partial \ln\sigma}{\partial\mu}, \quad e^2\kappa = \frac{\pi^2}{3} k_B^2 T\sigma, \qquad (13)$$

where σ is the electrical conductivity. Taking σ to vary linearly with carrier density, we find Q in the disordered limit takes on the same qualitative form as Eq.(3) in the clean limit. As a result, the qualitative features of energy-driven drag in the clean limit also appear in the disordered limit: namely, the avoided crossing of nodal lines (Fig. 3 (a)), a peak at zero doping, double sign reversal

along the diagonal $n_1 = n_2$ and a three-peak structure along the diagonal $n_1 = -n_2$ (Fig. 3 (b,c)). The T dependence of ρ_{21} is qualitatively similar in the diffusive and ballistic regimes. However, since the Wiedemann-Franz relation gives $\kappa \propto T$ (as opposed to $\kappa \propto 1/T$ in the ballistic regime), we find $\rho_{21} \propto T^2$ at lowest T and $\rho_{21} \propto T^{-4}$ at higher $T > \mu, \Delta$, as shown in Fig. 3 (d).

In summary, vertical energy transfer in graphene heterostructures has strong impact on lateral charge transport in the Coulomb drag regime, dominating the drag response at low doping. Drag measurements thus afford a unique probe of energy transfer at the nanoscale, a fundamental process which is not easily amenable to more conventional techniques such as calorimetry, and is key for the physics of strong interactions that occur near neutrality.

[1] L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M. I. Katsnelson, L. Eaves, S. V. Morozov, N. M. R. Peres, J. Leist, A. K. Geim, K. S. Novoselov, and L. A. Ponomarenko, Science 335 (6071), 947-950 (2012).

[2] W.-K. Tse and S. Das Sarma, Phys. Rev. B 75, 045333 (2007).

[3] B. N. Narozhny, Phys. Rev. B 76, 153409 (2007).

[4] R. Sensarma, E. H. Hwang, and S. Das Sarma, Phys. Rev. B 82, 195428 (2010).

[5] N. M. R. Peres, J. M. B. Lopes dos Santos, and A. H. Castro Neto, Europhys. Lett. 95, 18001 (2011).

[6] M. I. Katsnelson, Phys. Rev. B 84, 041407 (2011).

[7] B. N. Narozhny, M. Titov, I.V. Gornyi, and P.M. Ostrovsky, arXiv: 1110.6359 (2012)

[8] T. J. Gramila, J. P. Eisenstein, A. H. MacDonald, L. N. Pfeiffer, and K. W. West Phys. Rev. Lett. 66, 12161219 (1991)

[9] S. Kim, I. Jo, J. Nah, Z. Yao, S. K. Banerjee, and E. Tutuc, Phys. Rev. B 83, 161401(R) (2011).

[10] A. K. Geim, Electronic Properties of Graphene-BN Heterostructures, http://online.kitp.ucsb.edu/online/graphene_c12/geim/

[11] S. Das Sarma, E. H. Hwang, and W.-K. Tse, Phys. Rev. B, 75 121406 (2007).

[12] M. Müller, L. Fritz, and S. Sachdev, Phys. Rev. B., 78 115406 (2008)

[13] U. Sivan, P. M. Solomon, and H. Shtrikman Phys. Rev. Lett. 68, 11961199 (1992)

[14] J. Martin, N. Akerman, G. Ulbricht, T. Lohmann, J. H. Smet, K. Von Klitzing, A. Yacoby, Nat. Phys., 4, 144 (2008).

[15] Y. Zhang, V. W. Brar, C. Girit, A. Zettl, M. F. Crommie, Nature Physics 6, 722-726 (2009)

[16] R. Decker, Y. Wang, V. W. Brar, W. Regan, H. Tsai, Q. Wu, W. Gannett, A. Zettl, and M. F. Crommie, Nano Letters 11, 4631-4635 (2011)

[17] R. Bistritzer, A. H. MacDonald, Phys. Rev. Lett., **102** 206410 (2009).

[18] W.-K. Tse, S. Das Sarma, Phys. Rev. B, **79**, 235406 (2009).

[19] J. C. W. Song, M. Y. Reizer, and L. S. Levitov, arXiv: 1111.4678 (2011).

[20] see Appendix

[21] J.-H. Chen, C. Jang, S. Xiao, M. Ishigami, and M. S. Fuhrer, Nat.Nano. **3** 206 (2008).

[22] A. V. Andreev, S. A. Kivelson, and B. Spivak Phys. Rev. Lett., **106** 256804 (2011)

[23] M. Gibertini, A. Tomadin, F. Guinea, M. I. Katsnelson, and M. Polini, Phys. Rev. B , **85** 201405 (2012).

[24] B. Laikhtman and P. M. Solomon Phys. Rev. B 41, 9921-9929 (1990)

[25] A. O. Lyakhov, and E.G. Mischenko, Phys. Rev. B, **67** 041304 (2003).

APPENDIX A: HEAT CURRENT AND ONSAGER RECIPROCITY

Here we use Onsager reciprocity to obtain thermopower induced by a temperature imbalance in a Fermi gas. This can be done in a general form applicable in both the ballistic regime, when the mean free path is dominated by electron-electron scattering, and in the diffusive regime, when the mean free path is dominated by elastic scattering by disorder.

We start with recalling Onsager reciprocity for the heat and charge transport. Given charge current $\mathbf{j}$ and heat current $\mathbf{j}_q$ described by

$$-\mathbf{j} = L_{11}\frac{1}{T}\nabla\mu + L_{12}\nabla\frac{1}{T} \qquad (14)$$

$$\mathbf{j}_q = L_{21}\frac{1}{T}\nabla\mu + L_{22}\nabla\frac{1}{T}, \qquad (15)$$

the cross-couplings obey the Onsager relation $L_{12} = L_{21}$. Next, we consider the heat current, which can be obtained from the heat transport equation

$$\partial_t(C_{\text{el}}T) - \nabla\kappa\nabla T = -\mathbf{j}\cdot\nabla F[n(\mathbf{r}),T], \qquad (16)$$

where C_{el} is the electron heat capacity, $F[n(\mathbf{r}),T]$ is a function of the carrier density $n(\mathbf{r})$ to be determined later, T is the temperature and κ is the thermal conductivity. From Eq.(16) we find that the heat current is

$$\mathbf{j}_q = -\kappa\nabla T + \mathbf{j}F[n(\mathbf{r}),T] = -\kappa\nabla T - \sigma F[n(\mathbf{r}),T]\nabla\mu, \qquad (17)$$

which gives $L_{21} = -T\sigma F[n(\mathbf{r}),T]$ in Eq.(14). Here we have used $\nabla\cdot\mathbf{j} = 0$ and $\mathbf{j} = -\sigma\nabla\mu$, where $\sigma = L_{11}T$. Using the Onsager relation, the thermopower e.m.f. induced by the temperature gradient equals

$$\mathbf{E} = -\frac{F[n(\mathbf{r}),T]}{T}\nabla T. \qquad (18)$$

This result has general validity irrespective of the transport mechanism specifics, which are manifested through the form of $F[n(\mathbf{r}),T]$. This quantity equals the Seebeck coefficient s given by Eq.(13) in the diffusive regime, whereas in the ballistic regime it is given by Eq.(3).

The functional form of $F[n(\mathbf{r}),T]$ can be obtained by considering the kinetic equation (at steady state)

$$e\mathbf{E}\cdot\nabla_{\mathbf{p}}n(\mathbf{p},\mathbf{r}) = I_1 + I_2 \qquad (19)$$

where $e < 0$ is the carrier charge. Here we write the collision integral as a sum of momentum non-conserving and momentum conserving parts, $I_1 + I_2$, corresponding to disorder scattering and electron-electron scattering, respectively. Heat and change current can be expressed through a steady-states deviation from the equilibrium Fermi distribution, $\delta n = n - n_0$, as follows

$$\mathbf{j}_q = \sum_{\mathbf{p},i}(\epsilon_i - \mu)\mathbf{v}_{\mathbf{p},i}\delta n_i(\mathbf{p},\mathbf{r}), \quad \mathbf{j} = e\sum_{\mathbf{p},i}\mathbf{v}_{\mathbf{p},i}\delta n_i(\mathbf{p},\mathbf{r}). \qquad (20)$$

Here i labels the conduction and valence band states and $\mathbf{v}_{\mathbf{p},i} = \partial\epsilon_i/\partial\mathbf{p}$ and $\epsilon_i = \pm v|\mathbf{p}|$. Below we consider the diffusive and ballistic regimes. In the first case the mean free path is dominated by elastic momentum non-conserving scattering (I_1), in the second regime the mean free path is dominated by inelastic momentum-conserving scattering (I_2).

In the diffusive regime, neglecting I_2 and using the relaxation time approximation for I_1, we find

$$\delta n_{\mathbf{p}} = e\mathbf{E}\cdot\mathbf{v}_{\mathbf{p}}\tau(\epsilon)\partial n_F/\partial\epsilon, \qquad (21)$$

where τ describes elastic scattering by impurities. This gives the standard expressions for Seebeck and Pelteir coefficient described in Eq.(13) of the main text, so that $F = sT$.

In the ballistic regime, the fastest scattering mechanism comes from the (total) momentum conserving process of electron-electron scattering. As a result, we will neglect all other terms apart from I_2 in Eq. 19 and look for distributions, $n(\mathbf{p},\mathbf{r})$, that give a non-zero particle flow. At a nonzero total current, the non-equilibrium distribution can be written as $n_F(\epsilon - \mathbf{p}\cdot\mathbf{u})$, where the term $\mathbf{p}\cdot\mathbf{u}$ describes the change due to particle flow. This allows us to write δn as

$$\delta n_{\mathbf{p}i} = -\mathbf{u}\cdot\mathbf{p}\frac{\partial n_F}{\partial\epsilon} \qquad (22)$$

where the Fermi distribution, n_F, has a temperature that may depend on the flow and position. Using Eq. 22 and summing over both conduction and valence bands (where $\epsilon_c = v|\mathbf{p}|$ and $\epsilon_v = -v|\mathbf{p}|$, c and v refer to conduction and valence bands respectively) in Eq. 20 we obtain

$$\mathbf{j}_q = \frac{-\mathbf{u}}{2}\int_{-\infty}^{\infty}d\epsilon\,\nu(\epsilon)\epsilon(\epsilon - \mu)\frac{\partial n_F}{\partial\epsilon},$$

$$\mathbf{j} = \frac{-\mathbf{u}e}{2}\int_{-\infty}^{\infty}d\epsilon\,\nu(\epsilon)\epsilon\frac{\partial n_F}{\partial\epsilon} \qquad (23)$$

where $\nu(\epsilon) = 2|\epsilon|/(\pi\hbar^2 v^2)$ is the total density of states and we take into account that $v_{\mathbf{p}} \cdot \mathbf{p} = \epsilon$ with ϵ positive (negative) for the conduction (valence) band. Since the function $\epsilon\nu(\epsilon) \propto \epsilon|\epsilon|$, the integral **cannot** be evaluated for arbitrary ratio $k_B T/\mu$. Instead, we analyze the degenerate case, $\mu \gg k_B T$, in which case we use the Sommerfeld expansion to obtain

$$\mathbf{j} = \frac{e}{2}\mu\nu(\mu)\mathbf{u}, \quad \mathbf{j_q} = 2\zeta(2)\nu(\mu)(k_B T)^2 \mathbf{u} \qquad (24)$$

where we have used the identity $\int_0^\infty \frac{e^x x^2 dx}{(e^x+1)^2} = \zeta(2) = \pi^2/6$. Comparing both expressions to eliminate $\mathbf{u}$ we obtain

$$\mathbf{j_q} = 4\zeta(2)\frac{(k_B T)^2}{e\mu}\mathbf{j} \qquad (25)$$

The singularity at $\mu = 0$ is smeared by broadening of the Dirac point due to disorder and thermal fluctuations. We can account for smearing via

$$\mathbf{j_q} = \frac{4\zeta(2)}{e}\frac{(k_B T)^2\mu}{\mu^2 + \Delta^2(T)}\mathbf{j}, \quad \Delta^2(T) = \Delta_0^2 + \eta(k_B T)^2 \quad (26)$$

where Δ_0 is the Dirac point width parameter. This yields $F[n(\mathbf{r}), T] = Q$, giving Eq.(3) of the main text.

APPENDIX B: INTERLAYER ENERGY RELAXATION

In this Appendix, we derive the interlayer cooling power for the heterostructure system defined by the Hamiltonian in Eq.(6). Using Fermi's golden rule, the scattering rate is

$$W_{\mathbf{k_1}',\mathbf{k_1}} = \frac{2\pi N}{\hbar}\sum_{\substack{\mathbf{k_2},\mathbf{k_2}' \\ \mathbf{q}}} F_{\mathbf{k_2},\mathbf{k_2}'}|V_{\mathbf{q}}|^2 f(\epsilon_{\mathbf{k_2}})[1 - f(\epsilon_{\mathbf{k_2}'})]\delta_\epsilon \delta_1 \delta_2$$

$$(27)$$

where $\delta_\epsilon = \delta(\epsilon_{\mathbf{k_1}',\mathbf{k_1}} + \epsilon_{\mathbf{k_2}',\mathbf{k_2}})$, and $\delta_1 = \delta_{\mathbf{k_1}',\mathbf{k_1}+\mathbf{q}}$ and $\delta_2 = \delta_{\mathbf{k_2}',\mathbf{k_2}-\mathbf{q}}$. Here $N = 4$ is the number of spin/valley flavors, $\{1, 2\}$ denote the different layers, $\epsilon_{\mathbf{k}',\mathbf{k}} = \epsilon_{\mathbf{k}'} - \epsilon_{\mathbf{k}}$, and $F_{\mathbf{k},\mathbf{k}'} = |\langle \mathbf{k}'\alpha|\mathbf{k}\beta\rangle|^2$ is the coherence factor (α, β label states in the electron and hole Dirac cones). $V_{\mathbf{q}}$ is the screened inter-layer Coulomb interaction described below. The energy-loss power is

$$\mathcal{J} = -N\sum_{\mathbf{k_1},\mathbf{k_1}'} W_{\mathbf{k_1}',\mathbf{k_1}}(\epsilon_{\mathbf{k_1}'} - \epsilon_{\mathbf{k_1}})f(\epsilon_{\mathbf{k_1}})[1 - f(\epsilon_{\mathbf{k_1}'})]F_{\mathbf{k_1},\mathbf{k_1}'}$$

$$(28)$$

We can simplify the evaluation of these sums by writing $\delta(\epsilon_{\mathbf{k_1}',\mathbf{k_1}} + \epsilon_{\mathbf{k_2}',\mathbf{k_2}}) = \int_{-\infty}^\infty d\omega \delta(\epsilon_{\mathbf{k_1}',\mathbf{k_1}} - \omega)\delta(\omega + \epsilon_{\mathbf{k_2}',\mathbf{k_2}})$ and using the identity $f_s(\epsilon_{\mathbf{k}})[1 - f_s(\epsilon_{\mathbf{k}'})] = (f_s(\epsilon_{\mathbf{k}}) - f_s(\epsilon_{\mathbf{k}'})) \times (N_s(\epsilon_{\mathbf{k}',\mathbf{k}}) + 1)$ where $N(\omega) = 1/(e^{\omega/k_B T} - 1)$ is the Bose function taken at the electron temperature (of that particular layer) and $s = \{1, 2\}$ denotes the layers.

Using the quantities

$$\chi_s''(\mathbf{q}, \omega) = N\sum_{\mathbf{k}} F_{\mathbf{k},\mathbf{k}+\mathbf{q}}(f_s(\epsilon_{\mathbf{k}}) - f_s(\epsilon_{\mathbf{k}+\mathbf{q}}))\delta(\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}} - \omega).$$

$$(29)$$

where $\chi_s''(\mathbf{q}, \omega) = \frac{1}{\pi}\operatorname{Im}\Pi(\mathbf{q}, \omega)$ is the imaginary part of the susceptibility and $\Pi(\mathbf{q}, \omega) = N\sum_{\mathbf{k}} F_{\mathbf{k},\mathbf{k}+\mathbf{q}}\frac{f_s(\epsilon_{\mathbf{k}}) - f_s(\epsilon_{\mathbf{k}+\mathbf{q}})}{\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}} - \omega - i0}$ is the polarization operator. Using these we can re-write the energy loss power as

$$\mathcal{J} = \frac{\pi}{\hbar}\int d\omega\,\omega(N_2(\omega) - N_1(\omega))\sum_{\mathbf{q}}|V_{\mathbf{q}}|^2\chi_1''(\mathbf{q}, \omega)\chi_2''(\mathbf{q}, \omega)$$

$$(30)$$

where we have noted that $N(-\omega) = -(1 + N(\omega))$ and used the fact that $\chi_s''(\mathbf{q}, \omega) = -\chi_s''(-\mathbf{q}, -\omega)$ so as to only keep the odd part of the product $N_2(-\omega)(N_1(\omega) + 1)$.

The energy transfer between the two layers is dependent on the Coulomb interaction $V_{\mathbf{q}}$ between the layers. We treat $V_{\mathbf{q}}$ by accounting for polarization in both layers and screening in the RPA approximation. The RPA-screened coulomb interaction is

$$V_{\mathbf{q}} = \frac{V_{\mathbf{q}}^0}{1 - V_{\mathbf{q}}^0(\Pi_1(\mathbf{q}, \omega) + \Pi_2(\mathbf{q}, \omega))}, \quad V_{\mathbf{q}}^0 = \frac{2\pi e^2}{\varepsilon q} \quad (31)$$

where ε is the background dielectric constant. Here we ignored the correction due to finite interlayer spacing d, approximating the interlayer interaction $V_{\mathbf{q}} = \frac{2\pi e^2}{\varepsilon q}e^{-d|q|} \approx V_{\mathbf{q}}^0$. This approximation is valid when the layer separation d is small compared to the screening length in the layers, $dq_0 \ll 1$, $q_0 = -\Pi_1(\mathbf{q} = 0) - \Pi_2(\mathbf{q} = 0)$. We will be interested in the regime when d is small compared to the Fermi wavelength in the layers, for which the above approximation is adequate.

In the degenerate limit, $\mu \gg k_B T$, the polarization is given by

$$\Pi(q, \omega) = -\nu(\mu)\left(1 - \frac{\omega}{\sqrt{(\omega + i0)^2 - q^2 v^2}}\right) \qquad (32)$$

In the limit of $q \ll q_0$ where $q_0 = (2\pi e^2/\varepsilon)(\nu_1 + \nu_2)$, where $\nu_{1,2}$ is the density of states at the Fermi level in each layer, we can write $|V_q| = 1/|\Pi_1(\mathbf{q}, \omega) + \Pi_2(\mathbf{q}, \omega)|$. Noting that in the degenerate limit, intra-band transitions are the dominant processes we approximate $\chi''(\mathbf{q}, \omega)$ by the imaginary part of Eq.(32). This allows us to write the cooling power as

$$\mathcal{J}_{12} = \frac{\pi}{\hbar}\int d\omega\,\omega(N_2(\omega) - N_1(\omega))\sum_{|\mathbf{q}|>\omega/v}\frac{\nu_1\nu_2\omega^2}{(\nu_1 + \nu_2)^2 q^2 v^2}$$

$$(33)$$

where $T_0 = vq_0/k_B$. Using the identity $\int_{-\infty}^\infty \omega^3 d\omega N(\omega) = 2\cdot 3!\zeta(4)(k_B T)^4$, and recalling that the ultra-violet cutoff arises from screening via $q \ll q_0$ we obtain Eq.(7).